

An SNS-based literature review system for conducting a research survey

Yin, Chengjiu

Research Institute for Information Technology, Kyushu University

Hirokawa, Sachio

Research Institute for Information Technology, Kyushu University

Tabata, Yoshiyuki

Research Institute for Information Technology, Kyushu University

Yau, Jane Yin Kim

Dept of Computer Science, Malmö University

<https://hdl.handle.net/2324/1443973>

出版情報 : Proceedings of the 21st International Conference on Computers in Education, ICCE
2013, pp.404-410, 2013-01-01. UHAMKA PRESS

バージョン :

権利関係 :

An SNS-based Literature Review System for conducting a Research Survey

Chengjiu YIN^{a*}, Jane Yin-Kim YAU^b, Sachio HIROKAWA^c & Yoshiyuki TABATA^c

^a *Research Institute for Information Technology, Kyushu University, Japan*

^b *Dept of Computer Science, Malmö University, Sweden*

^c *Research Institute for Information Technology, Kyushu University, Japan*

*yin@kyushu-u.ac.jp

Abstract: It is necessary to perform a literature review before starting a new research project. However, many students do not know the procedures of performing a literature review. In this paper, based on the professional experiences and opinions of expert researchers, we describe an SNS-based literature review system to help students conduct research surveys. This system includes two search engines, one is an article search engine, which can help students conduct research surveys, and the other is a logging search engine, which allows students to learn from each other via their logs and share experience with other students. User models of the system as well as its functions are presented.

Keywords: Literature review, Learning by Searching, Data Mining, Search Engine

1. Introduction

Before initiating a new research project, it is necessary to conduct a literature review to be well aware of the current research state including substantive research findings of the area as well as theoretical, practical and methodological contributions (Yin et al., 2013a). As indicated by Hwang and Tsai (2011), the procedures of conducting research surveys allow educators and researchers to more consciously and have higher chances and possibilities to contribute to the relevant studies. "

A scientific literature survey should be accompanied with a document, which describes a research area and its related work and contains a number of relevant references. It is an important procedure for students to master in order to conduct research. Many researchers' supervisors may be consulted by their students seeking help on how to start a research project efficiently and eventually publish a paper.

Some of the supervisors may provide their students with some introductory literature. Using this, the students can perform a more detailed literature review. This makes it easier for the students to gain important articles, however, the ultimate problem is that they still do not know how and where to retrieve such relevant articles of importance. Hence, it is better to teach students how to find the important articles than to give the students articles directly. Moreover, the number of articles which are provided by supervisors may be limited, and some supervisors may request their students to read many journal papers and scholarly books according to their interests, and find some gaps in knowledge, which can be their research topic. However, the actual problem concerns how to find such valuable articles in their interests.

When starting to conduct research, the first step is to conduct a research survey. This paper focuses on the following three research problems/questions of conducting a research survey:

1. What kind of articles are important articles in the current research field?
2. What are the factors to determine these important articles?
3. How can one find important articles by using ICT and recommend these articles to students?

Regarding the first two questions, an interview was carried out by asking three research experts, who are working in Kyushu University. We asked them the following question: "Do you agree that the following factors – citation rating, journal ranking, survey or research paper, authors or founders - can be used to judge how valuable a paper is?" They answer the question with "Agree" or "Disagree".

Presented in table 1 are more information regarding these factors - 1) All three experts agreed that a paper with a high citation rating was a good paper; 2) Two of them agreed that the journal with a high impact factor was a good journal, as it undergoes a rigorous review process. The Impact Factor helps to evaluate a journal's relative importance. 3) A survey paper summarizes the current state of work in a particular field, which provides an overview of the main issues, results, and open questions in that field. On the other hand, research papers describe specific work (research, experimentation, or development of theory). Two of the experts agreed that high quality survey papers are very important. 4) In some fields, a number of people are considered the founders of that scientific field and in some other fields, they may even be considered as the "mother" or "father" of the field.¹ The founders' decision, technical knowledge, experience have significant contribution in their research areas. One of the experts confirmed that such papers belonging to a founder are very worthy of reading, as they normally have accumulated a lot of research experience and expertise in this area.

Table 1: Factors affecting the importance of research articles

Factors	Description
Citation rating	Papers with a high citation rating.
Journal ranking	Papers published in a journal with a high impact factor.
Survey/research papers	Survey papers provide an important overview of the main issues, results, and open questions in that field.
Authors/Founders	Papers published by certain well-known authors or founders of that scientific field are important.

In this paper, based on the results of the interview with the three experts, we propose an SNS-based literature review system for supporting students to perform literature research surveys. SNS is defined as a web-based service that allows individuals to: construct a profile within an organized framework; generate a list of other users with whom they share a connection; and navigate their own list of connections and view those made by others within the system (Boyd and Ellison, 2007; Yin, 2013).

This is a SNS system, which has two search engines- 1) Paper Search Engine (PSE), which is used to help students to conduct a research survey, and the other is a 2) Logs Search Engine (LSE), which allows students to learn from each other via their logs. The users will leave notes while they use the system. These logs are searching notes or history of their search processes that have been left by the expert researchers. Students maintain their learning logs, therefore a large number of logs on the Web will be accumulated. Our SNS-based system can analyze and share these learning logs by recording the users' search processes. Hence, as a result, students may become more easily aware of new knowledge, and learn search skills from others. By using the system, the students can learn how to obtain important articles, and a SNS member can also share their literature review methods and hence teach or learn from each other. This paper is organised as follows. In section 2, a literature review is presented. In section 3, the data preparation of creating the system search engine is presented. In section 4, a description of our system is explained. Finally, conclusions and future work are presented in section 5.

2. Literature Review

2.1 Learning by Searching (LBS) Strategy

The system focuses on providing a knowledge acquisition tool to effectively acquire knowledge and it employs a strategy called Learning by Searching – (LBS)', which we defined as the act of searching for information on the Web (Yin et al., 2013a). This is a cognitive process through which the user actively acquires knowledge. It is a strategy of learning and teaching that gives students opportunities to direct their own learning as they search on the Web.

The existing 'investigative learning' methods that use set teaching materials, such as textbooks and dictionaries, restrict the amount of knowledge that can be learned as they have prescribed educational contents and learning materials, such as using elementary school teaching materials (Murakami, 2010), students can get the right answers through searching activities.

¹ http://en.wikipedia.org/wiki/List_of_people_considered_father_or_mother_of_a_scientific_field

'LBS' is different from the existing 'investigative learning'. Compared with 'investigative learning', 'LBS' has the following characteristics:

- 1) Large amounts of Web data are available as target data.
- 2) Anyone can obtain Web data easily, however, the quality of information cannot be guaranteed.
- 3) Web data does not have correct answers, and therefore it is difficult to provide assessments.
- 4) Compare with existing teaching materials such as textbook, there is a wide variety of learning content on the web.
- 5) Different people may have different knowledge learning needs, with a high amount of individuality. In order to obtain the accurate information that matches their purpose, it is important to identify information sources which they would want to access.
- 6) While learning by searching, students will be able to become aware of the knowledge areas which they lack proficiency.
- 7) It is possible to generate new knowledge by using Knowledge Discovery and Data Mining technology to discovery large amounts of Web data.
- 8) It is possible to produce original ideas through Learning by Searching
- 9) It is possible to study through trial and error.

Other related works include: training students' web searching skills, such as: selecting keywords and/or phrases, the acquisition of search skills, and information gathering approach (Chiou, Hwang, & Tseng, 2009); observing and analyzing information-seeking behaviors in order to develop new environments that facilitate teachers' observation and analysis of the information-seeking behaviors of students in web-based learning environments (Tseng, Hwang, Tsai, & Tsai, 2009). This paper focuses on developing a LBS tool to help students to acquire knowledge effectively.

2.2 Classification of Literature Review

In general, a literature review can be classified in to 3 categories:

A) A survey of a specialized domain: This kind of survey is conducted by using textbooks, which can help to master the basic content, or obtain general/systematic knowledge of a specialized domain (Eom et al. 1993;). This kind of survey can be used to help grasp basic theory, technology, application, and data in the current research domain. It is suitable for beginners, general public, students and educators. This kind of research survey is conducted usually while writing a review paper.

B) Comparison with previous research: In order to determine the originality or effectiveness of research, one conducts a research survey to compare the research results with the previous research (Yin et al., 2013ab; Liu et al., 2011). This kind of research survey is conducted usually as a "Related work" section while writing a researcher paper.

C) Analyzing the research trend: In order to find some new research theme, one analyzes the research trend. Generally, journal papers and conference proceedings contain academic literature of higher quality, and these represent the latest developments in a specialized field (Wong & Monacob, 1995; Hwang & Tsai, 2011). Users may conduct research trends analysis with these documents.

Our system belongs to category C). We have built a " Milky Way Research Trend - MWRT" system for conducting surveys of scientific literature (Yin et al., 2013a), and the learning effectiveness of MWRT has been confirmed. Based on the MWRT, we proposed a SNS-based search engine to help analyze research trends. Not only can learners acquire the knowledge of their research fields, but also search and research skills by using the system.

3. Search Engines of the SNS-based System

As mentioned above, the core of the system is the Paper Search Engine and the Logs Search Engine. Data preparation is a key stage of creating these search engines. There are five steps of creating these search engines (Figure 1): 1) Data collection; 2) Stemming; 3) Creating a frequency file; 4) Creating a database; and 5) Building a system that enables keyword-based searches in the database.

Data preparation is conducted from step 1 to step 4. In these stages, “Stemmers”² and “GETA”³ were used to create the database. “Stemmers” were used to generate the frequency file as a textual index file. The frequency file is the internal data structure that stores the word index information; GETA was used to convert textual index file into a binary index file WAM (Word Article Matrix), which can be seen as a database. In sections 3.1 and 3.2, the data collection is described.

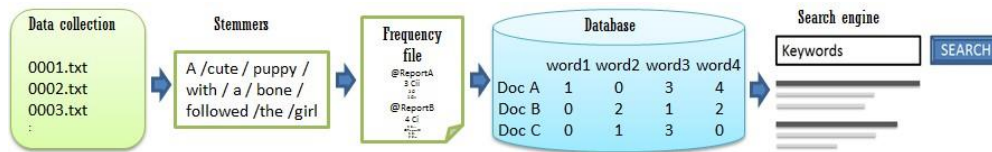


Figure 1. Five steps of creating a search engine

3.1 Paper Search Engine - Data Collection

With the development of the Internet, it is easy to obtain information via the Internet. There are many kinds of literature review sources, such as journal papers, conference papers, reports, and textbooks. Different sources have different characteristics. In order to refine accurate information and obtain data related to your needs, one must determine the right sources which match one’s purpose.

Published literature includes books, journals, proceedings, sci-tech reports, and patent literature. The characteristics of the published literature are as follows (Table 1):

Table 1: Comparison of published literature

	Latest progress	Detailed Data	Publish speed
Books	×	×	Slow
Journals	Δ	○	Fast
Proceedings	○	×	Very fast
Sci-Tech Report	○	○	Very fast

1. *Books*: Books require a longer period to write and publish; therefore, they are not suitable for keeping up-to-date with the latest progress of a particular field. Books are suitable for obtaining a general knowledge of a specialized domain, for mastering the basic content of a specialized problem or method in a short time, and for obtaining a preliminary understanding of the unfamiliar problem.

2. *Journals*: Journals focus on a specific discipline or field of study. Journals have the following characteristics: 1) high innovative content, 2) up-to-date reporting, and 3) contains timely reflection of domestic/international science and technology. In other words, journals are basic forms of scientific information transmission and exchange between academic researchers. Journals are also suitable for keeping up-to-date with the latest progress or for providing a deep understanding of a specialized field.

3. *Proceedings*: Generally, conference proceedings contain high-quality academic literature, and they represent the latest achievement/developments in a specialized field. Most of the proceedings are only presented by the results and it is not an inconvenience for knowing specific information. They are suitable for keeping up-to-date with the latest progress.

4. *Sci-Tech Report*: this refers to the official reports containing research results or progress during research studies issued by government or research departments. Sci-Tech reports are usually printed one year earlier than the journal. They report original information and results with detailed and reliable data. It is suitable for keeping up-to-date with the latest progress. It reflects the national and international trends and technology level.

In this paper, Journal and Proceeding are selected as data recourse of PSE, as they are suitable for keeping up-to-date with the latest progress. SciVerse Scopus (<http://www.info.sciverse.com/scopus/>) is

² <http://en.wikipedia.org/wiki/Stemming>

³ <http://geta.ex.nii.ac.jp/>

the world's largest abstract and citation database of peer-reviewed literature. We collected 13,326 “e-learning” articles and papers from Scopus and which contain journal and conference proceedings. Figure 2 shows the segments of search query.

3.2 Logs Search Engine - Data Collection

At the initial stage of the system, there is no community search logs data. However, we have created a search engine named “Milky Way”. Thirty-six students have used this “Milky Way” search engine to analyze the research trends of mobile learning articles. We used these students’ search logs as experiment data of the system. The search queries and browsing history of the experimental group were collected as the log data- 84 search queries and 192 browsed web page histories were collected. Each search queries contains the following:

- The number of users who have searched for the same information before
- Keywords of the query

Figure 3 shows the segments of log data. Each log data contains the following:

- IP address of the PC used by participants
- Date and time when the participants browsed the search results
- Paper number (1~13353)
- Keywords used by participants

```
...
22 mobile learning
17 mobile learning y:2010
9 mobile learning device y:2010
7 mobile learning y:2001
7 mobile learning c:Taiwan
...
```

Figure 2. Segment of search data

```
...
111.249.47.11,201201041255,4230,y:2011 mobile learn
133.5.7.108,201201041256, 5109,mobile learn ubiquitous
163.14.7.115,201201101200,8413,mobile learning c:Japan
163.14.7.82,201201101209,8523,mobile learning
...
```

Figure 3. Segment of log data.

4. A SNS-based Literature Review System - Description

After data have been prepared, we built a system which includes two search engines that enable keyword-based searches of the database. One is for searching research papers, named Papers Search Engine (PSE); the other is for search logs, named Logs Search Engine (LSE).

This system is a Web Learning system. Fan (2011) identified three benefits of Web Learning systems: 1) it can provide numerous and various learning materials; 2) it can retain the learners' autonomy; and 3) it is possible to facilitate collaborative learning between learners and teachers.

By using the system, users can share their search queries and browsing history. We refer to the search queries and browsing history as “survey history”. This system helps students to find the content they require. It also analyzes these community search logs, and shares the students’ searching knowledge and experiences with each other. Users can share problems or new knowledge and comment on others’ search activities.

4.1 Steps of using Paper Search Engine (PSE)

We design the PSE to analysis the collected papers and to select important papers. We used the factors in Table 1 to determine the importance of papers. We describe how to use the PSE to find important papers as follows (Figure 4):

Step 1, keywords selecting: at first, a user writes down a description in a few sentences that he wants to search, and then he selects some keywords from the content. Simultaneously, the system refines the chose keywords, such as recommending the professional or commonly used keywords and unifies the keywords with same meaning. For example, while a user entered the “mobile devices, learning” as a search query, the system will recommend the student to use the keywords “mobile learning”. If the user retrieved a small number of the articles, s/he can modify the keywords and retrieve again.

Step 2, refinement papers: the system recommends a batch of the relevant papers (about 100 papers), and then ranks the papers by the year of publication and the factors based on Table 1. The user refines/selects a certain number of papers, which can be seen as important papers (about 10 papers). The user should try to choose newly published papers to inform of new trends. The user also should select papers with higher number of pages in order to gather as much information as possible.

Step 3, expanding the references: the system lists the papers in the references of the selected papers, and then the user can find some more related/relevant papers (about 5 papers). The papers in the references have a higher correlation with the selected paper, so it is possible to find some important papers which the user requires.

Step 4, analyzing the retrieved results: by analyzing the papers, which are selected on Steps 2 and 3, the user can get important papers related to his search query. He can try to use other keywords or extend the scope of time to confirm if there are other important papers.

Step 5, read the valuable papers: the user can read important papers online, which were selected in Steps 2 and 3, and write down their reading memo on the system. These memos can be used to review and share with others. Through reading these papers, the user may identify his research direction. He can also write literature reviews by using the selected important papers and share it in the system.

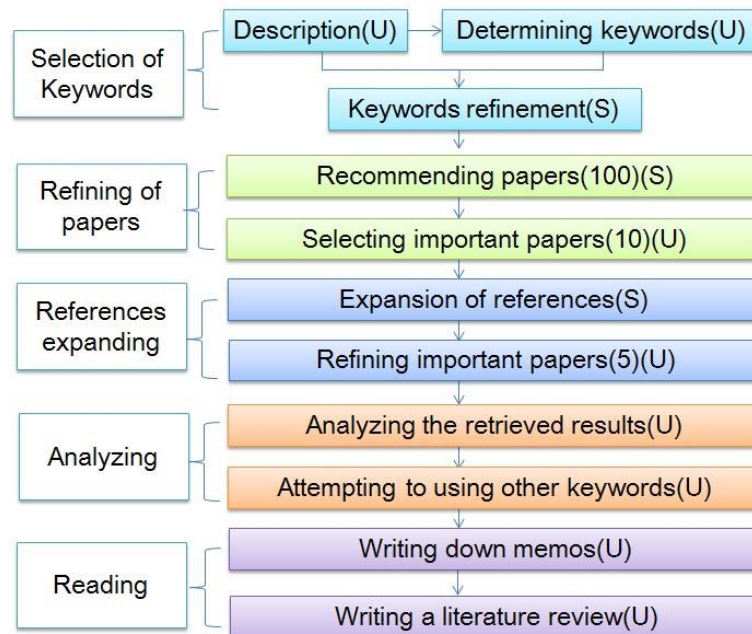


Figure 4. Step of using PSE.

4.2 User Model

The user model of the SNS-based system presents four kinds of users: “Expert”, “Novice”, “Active participant” and “Others” (Figure 5).

- A) Expert : An expert can use the PSE to do a research trend survey and leave searching memos. The search process will be recorded as survey history, which can be shared on the system. An expert can also answer the questions posed by other users.
- B) Novice: A novice can use the LSE to search experts’ survey history and/or searching memos, and learn research/searching skills /methods. Novices can also ask questions and discuss them with others in the community.
- C) Active Participant: An active participant can give novices some advice or comments and discuss with them any relating issues. Also, they can use the LSE to learn research/search skills..
- D) Others: this category of users may use the LSE to learn research/search skills, and may not participate in the discussion forums.



Figure 5. User model of the system.

4.3 System Functions

The system is composed of three parts: “Survey”, “Memo”, and “History”.

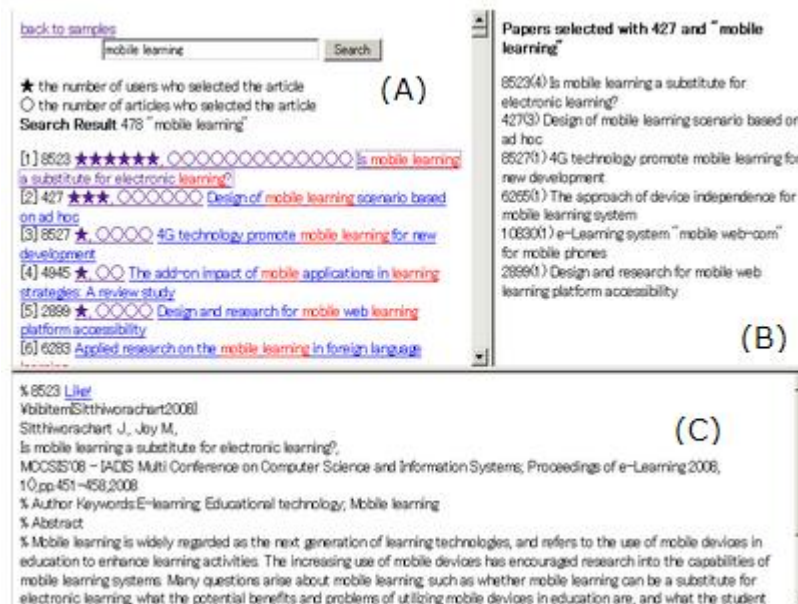


Figure 6. Interface of the system

1) “Survey”: The system provides survey functions such as search and classification. As shown in Figure 6 (A), the SNS system users can conduct keyword-based searches, and then the paper title and their numbers are displayed. When clicked ‘★’, ‘○’, or paper title, corresponding information will be displayed, as shown in Figure 6 (B).

‘★’ represents the users who selected the paper. When clicked ‘★’, the users who have browsed the paper are displayed in figure 6(B); ‘○’ represents the papers which were selected as related paper in

the past search results. When a user clicked the ‘○’, the number of times that the paper have been browsed is displayed in figure 6(B). When the paper title is clicked, the detail of the paper, such as authors and abstract is displayed in figure 6(C).

2)“Memo”: every user’s latest memo are displayed and the system also provides a forum for communicating and discussing about the memos.

3)“History”: Search keywords and survey history will be stored automatically.

4.4 Scenario of using the system

In this section, we give a general scenario for use with our system, as follows: A research team consisting of professors, senior and junior researchers were given two tasks: 1) to summarize the technology development process in mobile learning, and 2) to classify the main types of existing mobile learning applications.

A professor (expert) began to conduct the first task by using the PSE and inputted “mobile learning, technologies, trend” as search query items. Figure 7 shows the search results as “Android (2010), Smart Phones (2007), Wireless (2005), PDAs (2003)”; this means that these words have higher frequencies than others. In particular, ‘Android’ has been paid a lot of attention since 2010.

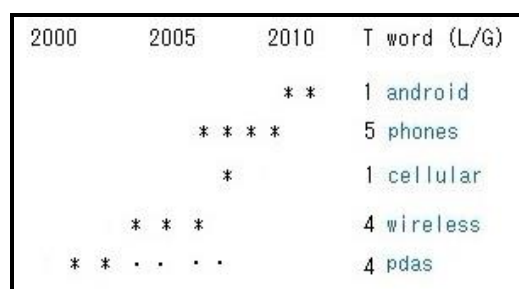


Figure 7. PSE search results

After that, the professor conducts the second task. According to his own past experience, he temporarily classified “mobile learning” into two categories: “Collaborative Data Gathering” and “Classroom Response Systems”. He inputs these two categories and clicks the “Classification” button. Simultaneously, he wrote down a memo as follows: “According to my own experience, I classified “mobile learning” into two categories”. Then the system refines the classification of mobile learning as these two categories. Thereafter, the system will recommend a new category called “Participatory Simulation”, as many researches classify mobile learning applications into three categories: “classroom response systems, participatory simulations, and collaborative data gathering” (Roschelle, 2003; Ogata et al., 2013). The survey process and memos are recorded automatically as survey history. The professor shares the survey history on the SNS-based system.

Senior and junior researchers use the LSE to search the survey histories of experts, and learn the research/search skills by browsing the experts’ survey history, as an example. They can discuss with each other through leaving their opinions and comments in the community forum.

5. Conclusion and Future Works

In this paper, we proposed a SNS-based system to help students conduct research surveys. It is very important to conduct research surveys for any academic research and may junior researchers may experience difficulties at the beginning of their research careers, as they are lack the experiences to locate appropriate keywords for searching within their research fields (Yin et al., 2013a).

Our system supports the analysis of research trends and offers learners opportunities to reflect their search behaviors. By browsing the experts’ survey history, students can learn research and search skills. In terms of pedagogical significance and effectiveness, our system has the following four features to facilitate these:

1) Users can use the system to analyze and compare knowledge/information and research methods in the literature review.

2) Opportunities are offered to users for reflection. When learners conduct a search, their search queries, memos, and browsing history are recorded, and therefore this helps learners to reflect their search behaviors.

3) Users' search skills and searching efficiency can be improved. By browsing the experts' survey history, students can master how to select search queries, conduct literature retrieval and analysis naturally, and therefore helping them to learn search skills.

4) SNS-based system users may share their searching history, so users may find the content they require easier.

In the future, we will develop the system and evaluate the learning effectiveness of the system.

Acknowledgements

This work was supported by JSPS KAKENHI Grant Number 25750084.

References

- Boyd, D. M., & Ellison, N. B. (2007), Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication*, 13(1), article 11. Available from: <http://jcmc.indiana.edu/vol13/issue1/boyd.ellison.html>. Retrieved May 20, 2013
- Chiou, C. K., Hwang, G. J., & Tseng, Judy C. R. (2009). An auto-scoring mechanism for evaluating problem-solving ability in a web-based learning environment. *Computers & Education*, 53(2), 261-272.
- Eom, S., Lee, M, Avaz, A.,(1993), Expert Systems Application Development Research in Business: A Selected Bibliography (1975-1989), *European Journal of Operational Research*, 68(1993), pp.278-290.
- Fan J., (2011). Constructing web-based learning environment for college English teaching. *Communications in Computer and Information Science*, 233 CCIS(PART 3), pp.515-521.
- Hwang , G. J. and Tsai, C.C. (2011), Research trends in mobile and ubiquitous learning: a review of publications in selected journals from 2001 to 2010. *British Journal of Educational Technology*, 42(4), pp. 65-70.
- Liu, Y., Yin, C., Ogata, H., Qiao, G. & Yano, Y.,(2011),A FAQ-Based e-Learning Environment to Support Japanese Language Learning, *International Journal of Distance Education Technologies*, 9(3), pp.45-55.
- Murakami, T. (2010), Development of New Scenario on Investigation Learning in Science Education, *Journal of educational research / Center for Educational Research and Training, Kyoto University of Education*, No. 10, 91-100.
- Ogata, H., Saito, N. A., Paredes J. R. G., San Martin, G. A., & Yano, Y. (2008). Supporting Classroom Activities with the BSUL System. *Educational Technology & Society*, 11 (1), 1-16.
- Roschelle, J. (2003) 'Unlocking the learning value of wireless mobile devices', *Journal of Computer Assisted Learning*, Vol. 19, pp.260-272.
- Tseng, Judy. C. R., Hwang, G. J., Tsai, P. S., & Tsai, C. C. (2009). Meta-analyzer: A web-based learning environment for analyzing student information searching behaviors. *International Journal of Innovative Computing, Information and Control*, 5(3), 567-579.
- Wong, B., Monacob, J., (1995), Expert system applications in business: A review and analysis of the literature (1977-1993), *Information & Management*, 29(3), pp. 141-152
- Yin, C., Hirokawa, S., Yau, J.Y., Nakatoh, T., Hashimoto, K. & Tabata, Y. (2013a), Analyzing Research Trends with Cross Tabulation Search Engine, *International Journal of Distance Education Technologies Special Issue on: "International Workshop on Technology-Enhanced Social Learning"*, 11(1), pp. 31- 44.
- Yin, C., Song, Y., Tabata, Y., Ogata, H., & Hwang, G.-J. (2013b). Developing and Implementing a Framework of Participatory Simulation for Mobile Learning Using Scaffolding. *Educational Technology & Society*, 16 (3), 137-150.